APPLICATION

in the name of

Patrick V. Warren and Ronald V. Swanson

of

Diversa

for

TRANSAMINASES AND AMINOTRANSFERASES

Lisa A. Haile, Ph.D.
FISH & RICHARDSON P.C.
4225 Executive Square, Suite 1400
La Jolla, CA 92037
(619) 678-5070 voice
(619) 678-5099 fax

I hereby certify under 37 CFR 1.10 that this correspondence is being deposited with the United States Postal Service as **Express Mail Post Office To Addressee** with sufficient postage on the date indicated below and is addressed to the Commissioner of Patents and Trademarks, Washington, D.C. 20231.

signature

AMY HUNT

name

DATE OF DEPOSIT: 4/21/98

Attorney Docket: 09010/016002

EXPRESS MAIL NO.: EM 153706597 US

TRANSAMINASES AND AMINOTRANSFERASES

a continuation of 09/069,226 filed 4/27/98, now US Pat 6,013,509 which is

This application is a Continuation of U.S. Patent Application No. 08/599,171 filed on February 9, 1996, now US Patent 5,814,473

3/19/02

This invention relates to newly identified polynucleotides, polypeptides encoded by such polynucleotides, the use of such polynucleotides and polypeptides, as well as the production and isolation of such polynucleotides and polypeptides. More particularly, the polynucleotides and polypeptides of the present invention have been putatively identified as transaminases and/or aminotransferases. Aminotransferases are enzymes that catalyze the transfer of amino groups from α-amino to α-keto acids. They are also called transaminases.

The α-amino groups of the 20 L-amino acids commonly found in proteins are removed during the oxidative degradation of the amino acids. The removal of the α-amino groups, the first step in the catabolism of most of the L-amino acids, is promoted by aminotransferases (or transaminases). In these transamination reactions, the α-amino group is transferred to the α-carbon atom of α-ketoglutarate, leaving behind the corresponding α-keto acid analog of the amino acid. There is no net deamination (*i.e.*, loss of amino groups) in such reactions because the α-ketoglutarate becomes aminated as the α-amino acid is deaminated. The effect of transamination reactions is to collect the amino groups from many different amino acids in the form of only one, namely, L-glutamate. The glutamate channels amino groups either into biosynthetic pathways or into a final sequence of reactions by which nitrogenous waste products are formed and then excreted.

Cells contain several different aminotransferases, many specific for α-ketoglutarate as the amino group acceptor. The aminotransferases differ in their specificity for the other substrate, the L-amino acid that donates the amino group, and

are named for the amino group donor. The reactions catalyzed by the aminotransferases are freely reversible, having an equilibrium constant of about 1.0 ($\Delta G^{0'} \simeq 0$ kJ/mol).

Aminotransferases are classic examples of enzymes catalyzing bimolecular ping-pong reactions. In such reactions the first substrate must leave the active site before the second substrate can bind. Thus the incoming amino acid binds to the active site, donates its amino group to pyridoxal phosphate, and departs in the form of an α-keto acid. Then the incoming α-keto acid is bound, accepts the amino group from pyridoxamine phosphate, and departs in the form of an amino acid.

The measurement of alanine aminotransferase and aspartate aminotransferase levels in blood serum is an important diagnostic procedure in medicine, used as an indicator of heart damage and to monitor recovery from the damage.

The polynucleotides and polypeptides of the present invention have been identified as transaminases and/or aminotransferases as a result of their enzymatic activity.

In accordance with one aspect of the present invention, there are provided novel enzymes, as well as active fragments, analogs and derivatives thereof.

In accordance with another aspect of the present invention, there are provided isolated nucleic acid molecules encoding the enzymes of the present invention including mRNAs, cDNAs, genomic DNAs as well as active analogs and fragments of such enzymes.

In accordance with another aspect of the present invention there are provided isolated nucleic acid molecules encoding mature polypeptides expressed by the DNA contained in ATCC Deposit No. _____.

In accordance with yet a further aspect of the present invention, there is provided a process for producing such polypeptides by recombinant techniques comprising culturing recombinant prokaryotic and/or eukaryotic host cells, containing a nucleic acid sequence of the present invention, under conditions promoting expression of said enzymes and subsequent recovery of said enzymes.

In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes, or polynucleotides encoding such enzymes for transferring an amino group from an α-amino acid to an α-keto acid. Most transaminases use L-amino acids as substrates, but as described below, it is also possible to convert the transaminases of the invention to use D-amino acids as substrates, thereby increasing their array of uses to include, for example, manufacture of synthetic pyrethroids and as components of β-lactam antibiotics. The transaminases of the invention are stable at high temperatures and in organic solvents and, thus, are superior for use with L- and/or D-amino acids for production of optically pure chiral compounds used in pharmaceutical, agricultural and other chemical industries.

In accordance with yet a further aspect of the present invention, there are also provided nucleic acid probes comprising nucleic acid molecules of sufficient length to hybridize to a nucleic acid sequence of the present invention.

In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes, or polynucleotides encoding such enzymes, for *in vitro* purposes related to scientific research, for example, to generate probes for identifying similar sequences which might encode similar enzymes from other organisms by using certain regions, *i.e.*, conserved sequence regions, of the nucleotide sequence.

These and other aspects of the present invention should be apparent to those skilled in the art from the teachings herein.

The following drawings are illustrative of embodiments of the invention and are not meant to limit the scope of the invention as encompassed by the claims.

Figure 1 is an illustration of the full-length DNA (SEQ ID NO:17) and corresponding deduced amino acid sequence (SEQ ID NO:25) of *Aquifex* aspartate transaminase A of the present invention. Sequencing was performed using a 378 automated DNA sequencer (Applied Biosystems, Inc.) for all sequences of the present invention.

Figure 2 is an illustration of the full-length DNA (SEQ ID NO:18) and corresponding deduced amino acid sequence (SEQ ID NO:26) of *Aquifex* aspartate aminotransferase B.

Figure 3 is an illustration of the full-length DNA (SEQ ID NO:19) and corresponding deduced amino acid sequence (SEQ ID NO:27) of *Aquifex* adenosyl-8-amino-7-oxononanoate aminotransferase.

Figure 4 is an illustration of the full-length DNA (SEQ ID NO:20) and corresponding deduced amino acid sequence (SEQ ID NO:28) of *Aquifex* acetylornithine aminotransferase.

Figure 5 is an illustration of the full-length DNA (SEQ ID NO:21) and corresponding deduced amino acid sequence (SEQ ID NO:29) of *Ammonifex degensii* aspartate aminotransferase.

Figure 6 is an illustration of the full-length DNA (SEQ ID NO:22) and corresponding deduced amino acid sequence (SEQ ID NO:30) of *Aquifex* glucosamine:fructose-6-phosphate aminotransferase.

Figure 7 is an illustration of the full-length DNA (SEQ ID NO:23) and corresponding deduced amino acid sequence (SEQ ID NO:31) of *Aquifex* histidinol-phosphate aminotransferase.

Figure 8 is an illustration of the full-length DNA (SEQ ID NO:24) and corresponding deduced amino acid sequence (SEQ ID NO:32) of *Pyrobaculum aerophilum* branched chain aminotransferase.

Figure 9 is a diagramatic illustration of the assay used to assess aminotransferase activity of the proteins using glutamate dehydrogenase.

The term "gene" means the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).

A coding sequence is "operably linked to" another coding sequence when RNA polymerase will transcribe the two coding sequences into a single mRNA, which is then translated into a single polypeptide having amino acids derived from both coding sequences. The coding sequences need not be contiguous to one another so long as the expressed sequences ultimately process to produce the desired protein.

"Recombinant" enzymes refer to enzymes produced by recombinant DNA techniques; *i.e.*, produced from cells transformed by an exogenous DNA construct encoding the desired enzyme. "Synthetic" enzymes are those prepared by chemical synthesis.

A DNA "coding sequence of" or a "nucleotide sequence encoding" a particular enzyme, is a DNA sequence which is transcribed and translated into an enzyme when placed under the control of appropriate regulatory sequences.

In accordance with an aspect of the present invention, there are provided isolated nucleic acids (polynucleotides) which encode for the mature enzymes having the deduced amino acid sequences of Figures 1-8 (SEQ ID NOS:17-32).

In accordance with another aspect of the present invention, there are provided isolated polynucleotides encoding the enzymes of the present invention. The deposited material is a mixture of genomic clones comprising DNA encoding an enzyme of the present invention. Each genomic clone comprising the respective DNA has been inserted into a pQE vector (Quiagen, Inc., Chatsworth, CA). The deposit has been deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852, USA, on December 13, 1995 and assigned ATCC Deposit No. _____.

The deposit(s) have been made under the terms of the Budapest Treaty on the International Recognition of the deposit of micro-organisms for purposes of patent procedure. The strains will be irrevocably and without restriction or condition released to the public upon the issuance of a patent. These deposits are provided merely as convenience to those of skill in the art and are not an admission that a deposit would be required under 35 U.S.C. §112. The sequences of the polynucleotides contained in the deposited materials, as well as the amino acid sequences of the polypeptides encoded thereby, are controlling in the event of any conflict with any description of sequences herein. A license may be required to make, use or sell the deposited materials, and no such license is hereby granted.

The polynucleotides of this invention were originally recovered from genomic DNA libraries derived from the following organisms:

Aquifex VF5 is a Eubacteria which was isolated in Vulcano, Italy. It is a gram-negative, rod-shaped, strictly chemolithoautotrophic, marine organism which grows

optimally at 85-90°C (T_{max}=95°C) at pH 6.8 in a high salt culture medium with Q as a substrate, and H_2/CO_2+0.5% O_2 in gas phase.

Ammonifex degensii KC4 is a new Eubacaterial organism isolated in Java, Indonesia. This Gram negative chemolithoautotroph has three respiration systems. The bacterium can utilize nitrate, sulfate, and sulfur. The organism grows optimally at 70°C, and pH 7.0, in a low salt culture medium with 0.2% nitrate as a substrate and H_2/CO_2 in gas phase.

Pyrobaculum aerophilium IM2 is a thermophilic sulfur archaea (Crenarchaeota) isolated in Ischia Maronti, Italy. It is a rod-shaped organism that grows optimally at 100°C at pH 7.0 in a low salt culture medium with nitrate, yeast extract, peptone, and O_2 as substrates and N_2/CO_2, O_2 in gas phase.

Accordingly, the polynucleotides and enzymes encoded thereby are identified by the organism from which they were isolated, and are sometimes hereinafter referred to as "VF5/ATA" (Figure 1 and SEQ ID NOS:17 and 25), "VF5/AAB" (Figure 2 and SEQ ID NOS:18 and 26), "VF5/A87A" (Figure 3 and SEQ ID NOS:19 and 27), "VF5/AOA" (Figure 4 and SEQ ID NOS:20 and 28), "KC4/AA" (Figure 5 and SEQ ID NOS:21 and 29), "VF5/GF6PA" (Figure 6 and SEQ ID NOS:22 and 30), "VF5/HPA" (Figure 7 and SEQ ID NOS:23 and 31) and "IM2/BCA" (Figure 8 and SEQ ID NOS:24 and 32).

The polynucleotides and polypeptides of the present invention show identity at the nucleotide and protein level to known genes and proteins encoded thereby as shown in Table 1.

Table 1

Enzyme	Gene w/closest Homology (Organism)	Protein Similarity (%)	Protein Identity (%)	DNA Identity (%)

VF5/ATA	*Bacillus subtilis*	57.5	38.3	50.1
VF5/AAB	*Sulfolobus solfataricus*	62.5	33.0	50.1
VF5/A87A	*Bacillus sphaericus BioA*	67.4	42.9	51
VF5/AOA	*Bacillus subtilis argD*	70.6	48.7	52.0
KC4/AA	*Bacillus YM-2 aspC*	72.6	52.7	52.0
VF5/GF6PA	*Rhizobium Leguminosarum NodM*	66.3	47.7	51.0
VF5/HPA	*Bacillus subtilis HisH/E.coli HisC (same gene)*	55.7	32.6	45.3
IM2/BCA	*E.coli iluE*	63.7	43.6	49.7

All the clones identified in Table 1 encode polypeptides which have transaminase or aminotransferase activity.

One means for isolating the nucleic acid molecules encoding the enzymes of the present invention is to probe a gene library with a natural or artificially designed probe using art recognized procedures (see, for example: Current Protocols in Molecular Biology, Ausubel F.M. *et al.* (EDS.) Green Publishing Company Assoc. and John Wiley Interscience, New York, 1989, 1992). It is appreciated by one skilled in the art that the polynucleotides of SEQ ID NOS:17-24, or fragments thereof (comprising at least 12 contiguous nucleotides), are particularly useful probes. Other particularly useful probes for this purpose are hybridizable fragments of the sequences of SEQ ID NOS:1-9 (*i.e.*, comprising at least 12 contiguous nucleotides).

With respect to nucleic acid sequences which hybridize to specific nucleic acid sequences disclosed herein, hybridization may be carried out under conditions of reduced stringency, medium stringency or even stringent conditions. As an example of oligonucleotide hybridization, a polymer membrane containing immobilized denatured nucleic acids is first prehybridized for 30 minutes at 45°C in a solution consisting of 0.9

M NaCl, 50 mM NaH_2PO_4, pH 7.0, 5.0 mM Na_2EDTA, 0.5% SDS, 10X Denhardt's, and 0.5 mg/mL polyriboadenylic acid. Approximately 2 X 10^7 cpm (specific activity 4-9 X 10^8 cpm/ug) of ^{32}P end-labeled oligonucleotide probe are then added to the solution. After 12-16 hours of incubation, the membrane is washed for 30 minutes at room temperature in 1X SET (150 mM NaCl, 20 mM Tris hydrochloride, pH 7.8, 1 mM Na_2EDTA) containing 0.5% SDS, followed by a 30 minute wash in fresh 1X SET at Tm -10°C (Tm is minus 10°C) for the oligo-nucleotide probe. The membrane is then exposed to auto-radiographic film for detection of hybridization signals.

Stringent conditions means hybridization will occur only if there is at least 90% identity, preferably at least 95% identity and most preferably at least 97% identity between the sequences. *See* J. Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual, 2d Ed.*, Cold Spring Harbor Laboratory (1989) which is hereby incorporated by reference in its entirety.

As used herein, a first DNA (RNA) sequence is at least 70% and preferably at least 80% identical to another DNA (RNA) sequence if there is at least 70% and preferably at least a 80% or 90% identity, respectively, between the bases of the first sequence and the bases of the another sequence, when properly aligned with each other, for example when aligned by BLASTN.

The present invention relates to polynucleotides which differ from the reference polynucleotide such that the changes are silent changes, for example the change does not or the changes do not alter the amino acid sequence encoded by the polynucleotide. The present invention also relates to nucleotide changes which result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference polynucleotide. In a preferred aspect of the invention these polypeptides retain the same biological action as the polypeptide encoded by the reference polynucleotide.

The polynucleotides of this invention were recovered from genomic gene libraries from the organisms listed in Table 1. Gene libraries were generated in the Lambda ZAP II cloning vector (Stratagene Cloning Systems). Mass excisions were performed on these libraries to generate libraries in the pBluescript phagemid. Libraries were generated and excisions were performed according to the protocols/methods hereinafter described.

The polynucleotides of the present invention may be in the form of RNA or DNA which DNA includes cDNA, genomic DNA, and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (anti-sense) strand. The coding sequences which encodes the mature enzymes may be identical to the coding sequences shown in Figures 1-8 (SEQ ID NOS:17-24) or may be a different coding sequence which coding sequence, as a result of the redundancy or degeneracy of the genetic code, encodes the same mature enzymes as the DNA of Figures 1-8 (SEQ ID NOS:17-24).

The polynucleotide which encodes for the mature enzyme of Figures 1-8 (SEQ ID NOS:25-32) may include, but is not limited to: only the coding sequence for the mature enzyme; the coding sequence for the mature enzyme and additional coding sequence such as a leader sequence or a proprotein sequence; the coding sequence for the mature enzyme (and optionally additional coding sequence) and non-coding sequence, such as introns or non-coding sequence 5' and/or 3' of the coding sequence for the mature enzyme.

Thus, the term "polynucleotide encoding an enzyme (protein)" encompasses a polynucleotide which includes only coding sequence for the enzyme as well as a polynucleotide which includes additional coding and/or non-coding sequence.

The present invention further relates to variants of the hereinabove described polynucleotides which encode for fragments, analogs and derivatives of the enzymes

having the deduced amino acid sequences of Figures 1-8 (SEQ ID NOS:25-32). The variant of the polynucleotide may be a naturally occurring allelic variant of the polynucleotide or a non-naturally occurring variant of the polynucleotide.

Thus, the present invention includes polynucleotides encoding the same mature enzymes as shown in Figures 1-8 (SEQ ID NOS:17-24) as well as variants of such polynucleotides which variants encode for a fragment, derivative or analog of the enzymes of Figures 1-8 (SEQ ID NOS:17-24). Such nucleotide variants include deletion variants, substitution variants and addition or insertion variants.

As hereinabove indicated, the polynucleotides may have a coding sequence which is a naturally occurring allelic variant of the coding sequences shown in Figures 1-8 (SEQ ID NOS:17-24). As known in the art, an allelic variant is an alternate form of a polynucleotide sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded enzyme. Also, using directed and other evolution strategies, one may make very minor changes in DNA sequence which can result in major changes in function.

Fragments of the full length gene of the present invention may be used as hybridization probes for a cDNA or a genomic library to isolate the full length DNA and to isolate other DNAs which have a high sequence similarity to the gene or similar biological activity. Probes of this type preferably have at least 10, preferably at least 15, and even more preferably at least 30 bases and may contain, for example, at least 50 or more bases. The probe may also be used to identify a DNA clone corresponding to a full length transcript and a genomic clone or clones that contain the complete gene including regulatory and promotor regions, exons and introns. An example of a screen comprises isolating the coding region of the gene by using the known DNA sequence to synthesize an oligonucleotide probe. Labeled oligonucleotides having a sequence complementary or identical to that of the gene or portion of the gene sequences of the

present invention are used to screen a library of genomic DNA to determine which members of the library the probe hybridizes to.

It is also appreciated that such probes can be and are preferably labeled with an analytically detectable reagent to facilitate identification of the probe. Useful reagents include but are not limited to radioactivity, fluorescent dyes or enzymes capable of catalyzing the formation of a detectable product. The probes are thus useful to isolate complementary copies of DNA from other sources or to screen such sources for related sequences.

The present invention further relates to polynucleotides which hybridize to the hereinabove-described sequences if there is at least 70%, preferably at least 90%, and more preferably at least 95% identity between the sequences. The present invention particularly relates to polynucleotides which hybridize under stringent conditions to the hereinabove-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. The polynucleotides which hybridize to the hereinabove described polynucleotides in a preferred embodiment encode enzymes which either retain substantially the same biological function or activity as the mature enzyme encoded by the DNA of Figures 1-8 (SEQ ID NOS:17-24).

Alternatively, the polynucleotide may have at least 15 bases, preferably at least 30 bases, and more preferably at least 50 bases which hybridize to any part of a polynucleotide of the present invention and which has an identity thereto, as hereinabove described, and which may or may not retain activity. For example, such polynucleotides may be employed as probes for the polynucleotides of SEQ ID NOS:17-24, for example, for recovery of the polynucleotide or as a diagnostic probe or as a PCR primer.

Thus, the present invention is directed to polynucleotides having at least a 70% identity, preferably at least 90% identity and more preferably at least a 95% identity to a polynucleotide which encodes the enzymes of SEQ ID NOS:25-32 as well as fragments thereof, which fragments have at least 15 bases, preferably at least 30 bases and most preferably at least 50 bases, which fragments are at least 90% identical, preferably at least 95% identical and most preferably at least 97% identical under stringent conditions to any portion of a polynucleotide of the present invention.

The present invention further relates to enzymes which have the deduced amino acid sequences of Figures 1-8 (SEQ ID NOS:17-24) as well as fragments, analogs and derivatives of such enzyme.

The terms "fragment," "derivative" and "analog" when referring to the enzymes of Figures 1-8 (SEQ ID NOS:25-32) means enzymes which retain essentially the same biological function or activity as such enzymes. Thus, an analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature enzyme.

The enzymes of the present invention may be a recombinant enzyme, a natural enzyme or a synthetic enzyme, preferably a recombinant enzyme.

The fragment, derivative or analog of the enzymes of Figures 1-8 (SEQ ID NOS:25-32) may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature enzyme is fused with another compound, such as a compound to increase the half-life of the enzyme (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature enzyme, such as a leader or secretory sequence or a sequence which is

employed for purification of the mature enzyme or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

The enzymes and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or enzyme present in a living animal is not isolated, but the same polynucleotide or enzyme, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or enzymes could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.

The enzymes of the present invention include the enzymes of SEQ ID NOS:25-32 (in particular the mature enzyme) as well as enzymes which have at least 70% similarity (preferably at least 70% identity) to the enzymes of SEQ ID NOS:25-32 and more preferably at least 90% similarity (more preferably at least 90% identity) to the enzymes of SEQ ID NOS:25-32 and still more preferably at least 95% similarity (still more preferably at least 95% identity) to the enzymes of SEQ ID NOS:25-32 and also include portions of such enzymes with such portion of the enzyme generally containing at least 30 amino acids and more preferably at least 50 amino acids.

As known in the art "similarity" between two enzymes is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one enzyme to the sequence of a second enzyme.

A variant, *i.e.* a "fragment", "analog" or "derivative" polypeptide, and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions, fusions and truncations, which may be present in any combination.

Among preferred variants are those that vary from a reference by conservative amino acid substitutions. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr.

Most highly preferred are variants which retain the same biological function and activity as the reference polypeptide from which it varies.

Fragments or portions of the enzymes of the present invention may be employed for producing the corresponding full-length enzyme by peptide synthesis; therefore, the fragments may be employed as intermediates for producing the full-length enzymes. Fragments or portions of the polynucleotides of the present invention may be used to synthesize full-length polynucleotides of the present invention.

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of enzymes of the invention by recombinant techniques.

Host cells are genetically engineered (transduced or transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector such as an expression vector. The vector may be, for example, in the form of a plasmid, a phage, *etc.* The engineered host cells can be cultured in conventional nutrient media

modified as appropriate for activating promoters, selecting transformants or amplifying the genes of the present invention. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

The polynucleotides of the present invention may be employed for producing enzymes by recombinant techniques. Thus, for example, the polynucleotide may be included in any one of a variety of expression vectors for expressing an enzyme. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, *e.g.*, derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host.

The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. As representative examples of such promoters, there may be mentioned: LTR or SV40 promoter, the *E. coli*. *lac* or *trp*, the phage lambda P_L promoter and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such

as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E. coli*.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as *E. coli, Streptomyces, Bacillus subtilis*; fungal cells, such as yeast; insect cells such as *Drosophila S2* and *Spodoptera Sf9*; animal cells such as CHO, COS or Bowes melanoma; adenoviruses; plant cells, *etc.* The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

More particularly, the present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example; Bacterial: pQE70, pQE60, pQE-9 (Qiagen), pBluescript II KS, ptrc99a, pKK223-3, pDR540, pRIT2T (Pharmacia); Eukaryotic: pXT1, pSG5 (Stratagene) pSVK3, pBPV, pMSG, pSVL SV40 (Pharmacia). However, any other plasmid or vector may be used as long as they are replicable and viable in the host.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters

include lacI, lacZ, T3, T7, gpt, lambda P_R, P_L and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In a further embodiment, the present invention relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (Davis, L., Dibner, M., Battey, I., *Basic Methods in Molecular Biology*, (1986)).

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the enzymes of the invention can be synthetically produced by conventional peptide synthesizers.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook *et al., Molecular Cloning: A Laboratory Manual, Second Edition*, Cold Spring Harbor, N.Y., (1989), the disclosure of which is hereby incorporated by reference.

Transcription of the DNA encoding the enzymes of the present invention by higher eukaryotes is increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act on a promoter to increase its transcription. Examples include the SV40 enhancer on the late side of the replication origin bp 100 to 270, a cytomegalovirus early promoter enhancer,

the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), α-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated enzyme. Optionally, the heterologous sequence can encode a fusion enzyme including an N-terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli, Bacillus subtilis, Salmonella typhimurium* and various species within the genera Pseudomonas, Streptomyces, and Staphylococcus, although others may also be employed as a matter of choice.

As a representative but nonlimiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example,

pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and pGEM1 (Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well known to those skilled in the art.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, *Cell, 23*:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

The enzyme can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic

interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

The enzymes of the present invention may be a naturally purified product, or a product of chemical synthetic procedures, or produced by recombinant techniques from a prokaryotic or eukaryotic host (for example, by bacterial, yeast, higher plant, insect and mammalian cells in culture). Depending upon the host employed in a recombinant production procedure, the enzymes of the present invention may be glycosylated or may be non-glycosylated. Enzymes of the invention may or may not also include an initial methionine amino acid residue.

Transaminases are a group of key enzymes in the metabolism of amino acids and amino sugars and are found in all organisms from microbes to mammals. In the transamination reaction, an amino group is transferred from an amino acid to an α-keto acid. Pyridoxal phosphate is required as a co-factor to mediate the transfer of the amino group without liberation of ammonia.

Amino acids currently have applications as additives to aminal feed, human nutritional supplements, components in infusion solutions, and synthetic intermediates for manufacture of pharmaceuticals and agricultural products. For example, L-glutamic acid is best known as a flavor enhancer for human food. L-lysine and L-methionine are large volume additives to animal feed and human supplements. L-tryptophan and L-threonine have similar potential applications. L-phenylalanine and L-aspartic acid have very important market potential as key components in the manufacture of the low-calorie sweetener aspartame, and other promising low-calorie sweeteners have compositions containing certain amino acids as well. Infusion solutions require a large range of amino acids including those essential ones in human diets.

Transaminases are highly stereoselective, and most use L-amino acids as substrates. Using the approach disclosed in a commonly assigned, copending provisional application Serial No. 60/008,316, filed on December 7, 1995 and entitled "Combinatorial Enzyme Development," the disclosure of which is incorporated herein by reference in its entirety, one can convert the transaminases of the invention to use D-amino acids as substrates. Such conversion makes possible a broader array of transaminase applications. For instance, D-valine can be used in the manufacture of synthetic pyrethroids. D-phenylglycine and its derivatives can be useful as components of β-lactam antibiotics. Further, the thermostable transaminases have superior stability at higher temperatures and in organic solvents. Thus, they are better suited to utilize either L- and/or D-amino acids for production of optically pure chiral compounds used in pharmaceutical, agricultural, and other chemical manufactures.

There are a number of reasons to employ transaminases in industrial-scale production of amino acids and their derivatives.

1) Transaminases can catalyze stereoselective synthesis of D- or L-amino acids from their corresponding α-keto acids. Therefore no L- or D-isomers are produced, and no resolution is required.

2) Transaminases have uniformly high catalytic rates, capable of converting up to 400 μmoles of substrates per minute per mg enzyme.

3) Many required α-keto acids can be conveniently prepared by chemical synthesis at low cost.

4) The capital investment for an immobilized enzyme process using transaminases is much lower than for a large scale fermentation process, and productivity of the bioreactor is often an order of magnitude higher.

5) The technology is generally applicable to a broad range of D- or L-amino acids because transaminases exist with varying specificities. Such broad scope allows a number of different L- or D-amino acids to be produced with the same equipment and often the same biocatalyst.

Antibodies generated against the enzymes corresponding to a sequence of the present invention can be obtained by direct injection of the enzymes into an animal or by administering the enzymes to an animal, preferably a nonhuman. The antibody so obtained will then bind the enzymes itself. In this manner, even a sequence encoding only a fragment of the enzymes can be used to generate antibodies binding the whole native enzymes. Such antibodies can then be used to isolate the enzyme from cells expressing that enzyme.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, *Nature, 256*:495-497, 1975), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunology Today 4*:72, 1983), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole *et al.*, in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96, 1985).

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to immunogenic enzyme products of this invention. Also, transgenic mice may be used to express humanized antibodies to immunogenic enzyme products of this invention.

Antibodies generated against an enzyme of the present invention may be used in screening for similar enzymes from other organisms and samples. Such screening techniques are known in the art, for example, one such screening assay is described in Sambrook and Maniatis, Molecular Cloning: A Laboratory Manual (2d Ed.), vol.

2:Section 8.49, Cold Spring Harbor Laboratory, 1989, which is hereby incorporated by reference in its entirety.

The present invention will be further described with reference to the following examples; however, it is to be understood that the present invention is not limited to such examples. All parts or amounts, unless otherwise specified, are by weight.

In order to facilitate understanding of the following examples certain frequently occurring methods and/or terms will be described.

"Plasmids" are designated by a lower case "p" preceded and/or followed by capital letters and/or numbers. The starting plasmids herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in accord with published procedures. In addition, equivalent plasmids to those described are known in the art and will be apparent to the ordinarily skilled artisan.

"Digestion" of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors and other requirements were used as would be known to the ordinarily skilled artisan. For analytical purposes, typically 1 μg of plasmid or DNA fragment is used with about 2 units of enzyme in about 20 μl of buffer solution. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μg of DNA are digested with 20 to 250 units of enzyme in a larger volume. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37°C are ordinarily used, but may vary in accordance with the supplier's instructions. After digestion the reaction is electrophoresed directly on a polyacrylamide gel to isolate the desired fragment.

Size separation of the cleaved fragments is performed using 8 percent polyacrylamide gel described by Goeddel *et al.*, *Nucleic Acids Res.*, *8*:4057 (1980).

"Oligonucleotides" refers to either a single stranded polydeoxynucleotide or two complementary polydeoxynucleotide strands which may be chemically synthesized. Such synthetic oligonucleotides have no 5' phosphate and thus will not ligate to another oligonucleotide without adding a phosphate with an ATP in the presence of a kinase. A synthetic oligonucleotide will ligate to a fragment that has not been dephosphorylated.

"Ligation" refers to the process of forming phosphodiester bonds between two double stranded nucleic acid fragments (Maniatis, T., *et al.*, *Id.*, p. 146). Unless otherwise provided, ligation may be accomplished using known buffers and conditions with 10 units of T4 DNA ligase ("ligase") per 0.5 μg of approximately equimolar amounts of the DNA fragments to be ligated.

Unless otherwise stated, transformation was performed as described in Sambrook and Maniatis, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, 1989.

Example 1

Bacterial Expression and Purification of Transaminases and Aminotransferases

DNA encoding the enzymes of the present invention, SEQ ID NOS:25 through 32, were initially amplified from a pBluescript vector containing the DNA by the PCR technique using the primers noted herein. The amplified sequences were then inserted into the respective PQE vector listed beneath the primer sequences, and the enzyme was expressed according to the protocols set forth herein. The genomic DNA has also been used as a template for the PCR amplification, *i.e.*, once a positive clone has been identified and primer sequences determined using the cDNA, it was then possible to return to the genomic DNA and directly amplify the desired sequence(s) there. The 5' and 3' primer sequences and the vector for the respective genes are as follows:

***Aquifex* Aspartate Transaminase A**

aspa501 5′ CCGAGAATTCATTAAAGAGGAGAAATTAACTATGATTGAAGACCCTATGGAC (SEQ. ID NO:1)

aspa301 3′ CGAAGATCTTTAGCACTTCTCTCAGGTTC (SEQ. ID NO:2)

vector: pQET1

***Aquifex* Aspartate Aminotransferase B**

aspb501 5′ CCGAGAATTCATTAAAGAGGAGAAATTAACTATGGACAGGCTTGAAAAAGTA (SEQ ID NO:3)

aspb301 3′ CGGAAGATCTTCAGCTAAGCTTCTCTAAGAA (SEQ ID NO:4)

vector: pQET1

***Aquifex* Adenosyl-8-amino-7-oxononanoate Aminotransferase**

ameth501 5′ CCGACAATTGATTAAAGAGGAGAAATTAACTATGTGGGAATTAGACCCTAAA (SEQ ID NO:5)

ameth301 3′ CGGAGGATCCCTACACCTCTTTTTCAAGCT (SEQ ID NO:6)

vector: pQET12

***Aquifex* Acetylornithine Aminotransferase**

aorn 501 5′ CCGACAATTGATTAAAGAGGAGAAATTAACTATGACATACTTAATGAACAAT (SEQ ID NO:7)

aorn 301 3′ CGGAAGATCTTTATGAGAAGTCCCTTTCAAG (SEQ ID NO:8)

vector: pQET12

***Ammonifex degensii* Aspartate Aminotransferase**

adasp 501 5′ CCGAGAATTCATTAAAGAGGAGAAATTAACTATGCGGAAACTGGCCGAGCGG (SEQ ID NO:9)

adasp 301 3′ CGGAGGATCCTTAAAGTGCCGCTTCGATCAA (SEQ ID NO:10)

vector: pQET12

***Aquifex* Glucosamine:Fructose-6-phosphate Aminotransferase**

glut 501 5′ CCGACAATTGATTAAAGAGGAGAAATTAACTATGTGCGGGATAGTCGGATAC (SEQ ID NO:11)

glut 301 3′ CGGAAGATCTTTATTCCACCGTGACCGTTTT (SEQ ID NO:12)

vector: pQET1

***Aquifex* Histadine-phosphate Aminotransferase**

his 501 5′ CCGACAATTGATTAAAGAGGAGAAATTAACTATGATACCCCAGAGGATTAAG (SEQ ID NO:13)
his 301 3′ CGGAAGATCTTTAAAGAGAGCTTGAAAGGGA (SEQ ID NO:14)
vector: pQET1

***Pyrobacullum aerophilum* Branched Chain Aminotransferase**

bcat 501 5′ CCGAGAATTCATTAAAGAGGAGAAATTAACTATGAAGCCGTACGCTAAATAT (SEQ ID NO:15)
bcat 301 3′ CGGAAGATCTCTAATACACAGGAGTGATCCA (SEQ ID NO:16)
vector: pQET1

The restriction enzyme sites indicated correspond to the restriction enzyme sites on the bacterial expression vector indicated for the respective gene (Qiagen, Inc. Chatsworth, CA). The pQE vector encodes antibiotic resistance (Amp^r), a bacterial origin of replication (ori), an IPTG-regulatable promoter operator (P/O), a ribosome binding site (RBS), a 6-His tag and restriction enzyme sites.

The pQE vector was digested with the restriction enzymes indicated. The amplified sequences were ligated into the respective pQE vector and inserted in frame with the sequence encoding for the RBS. The ligation mixture was then used to transform the *E. coli* strain M15/pREP4 (Qiagen, Inc.) by electroporation. M15/pREP4 contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan^r). Transformants were identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies were selected. Plasmid DNA was isolated and confirmed by restriction analysis. Clones containing the desired constructs were grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture was used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells were grown to an optical density 600 ($O.D._{600}$) of between 0.4 and 0.6. IPTG ("Isopropyl-B-D-thiogalacto pyranoside") was then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression. Cells were grown an extra 3 to 4 hours. Cells were then harvested by centrifugation.

The primer sequences set out above may also be employed to isolate the target gene from the deposited material by hybridization techniques described above.

Example 2

Isolation of a Selected Clone from the Deposited Genomic Clones

The two oligonucleotide primers corresponding to the gene of interest are used to amplify the gene from the deposited material. A polymerase chain reaction is carried out in 25 μl of reaction mixture with 0.1 μg of the DNA of the gene of interest. The reaction mixture is 1.5-5 mM $MgCl_2$, 0.01% (w/v) gelatin, 20 μM each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 1.25 Unit of Taq polymerase. Thirty cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with the Perkin-Elmer Cetus 9600 thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the gene of interest by subcloning and sequencing the DNA product.

Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, within the scope of the appended claims, the invention may be practiced otherwise than as particularly described.